1X

УДК 004

Т.С. Хашан
Институт прикладной математики и механики НАН Украины, г. Донецк
khashan@mail.ru

Автоматическое определение границ текущих объектов исследования для решения задач пеленгации, сегментации речевых и специальных сигналов

Разработан метод автоматического определения границ текущих объектов исследования в образах звуковых волн, инвариантный к темпу, тембру и интенсивности принятого акустического сигнала. В качестве текущих объектов исследования выступают отрезки амиплитудно-временного представления волнового процесса, соответствующие фонологическим единицам, в формировании которых участвует голосовой источник (для речевых сигналов) и тональным сигналам (для специальных сигналов). На основе предлагаемого метода разработан алгоритм, который успешно применен в задачах пеленгации на макетной установке системы технического слуха.

Введение

В проблеме искусственного интеллекта, которая является одной из ключевых в робототехнике, важное место занимают вопросы создания технических аналогов органов чувств. Их действие основано на моделировании операций, выполняемых органами чувств живых организмов. К их числу относятся: зрение, слух, осязание, вкус и обоняние. Акустические средства коммуникации имеют ряд преимуществ перед всеми остальными средствами: скоростью, гибкостью (проникают через не очень плотные вещества, огибают препятствия), не зависят от светового фактора, всенаправленного действия [1]. В настоящее время актуальной является разработка алгоритмов обработки акустических сигналов для интеллектуальных систем технического слуха (СТС): пеленгация речевых и специальных сигналов – свистка, дудочки (далее акустических сигналов – АС); получение и обработка «звуковой картины» окружающего пространства, с целью обнаружения требуемых АС и их верификация.

Качественное решение указанных задач зависит от метода выбора границ текущего объекта исследования (ОИ) – временного интервала, на котором определяются искомые характеристики сигнала. Например, для решения задач пеленгации АС в качестве ОИ выбирают некоторый временной отрезок, на котором при помощи частотных методов или методов сравнения принятого микрофонами сигнала оценивается временная задержка или фазовый сдвиг. Границы текущего ОИ для каждого типа АС определяются субъективно на основе статистических данных, и их выбор влияет на точность пеленгации [2-4]. Очевидно, что решение задачи автоматического определения границ текущих ОИ безотносительно типа пеленгуемого сигнала позволит повысить точность пеленгации.

Следует отметить, что в АС, указанных выше, возможно различить звуки определенной высоты (звуки подавляющего большинства музыкальных инструментов, также пение), а высота звука (или высота тона) определяется частотой собственного колебания системы, а следовательно, самой природой этой системы (для речи – час-

тота вибрации голосовых связок), поэтому в качестве ОИ должны выступать отрезки сигнала, соответствующие функционированию голосового источника (для речи) или тональным звукам (для специальных сигналов).

Решение задачи автоматического определения границ текущих ОИ также возникает при сегментации речевых сигналов (РС), когда необходимо определить границы сегментов, соответствующих фонологическим единицам, в формировании которых участвует голосовой источник (носитель тона). Носителями тона чаще всего являются гласные, но встречаются языки, где в этой роли могут выступать и согласные, чаще всего сонанты. Для решения задачи сегментации тональных звуков используют «окна» фиксированной длительности, их размер и границы выбираются на основе субъективных данных [5-8]. Очевидно, что вариативность произнесения одного слова у одного диктора бесконечное множество, поскольку предопределяется его темпом, тембром, интенсивностью, а также рядом психологических факторов. Поэтому априорное задание границ и длины ОИ без учета случайности поведения характеристик РС не позволяет качественно решить указанную задачу.

Общность целей при разработке алгоритмов пеленгации и сегментации РС позволяет в дальнейшем перейти к следующей постановке задачи.

Постановка задачи

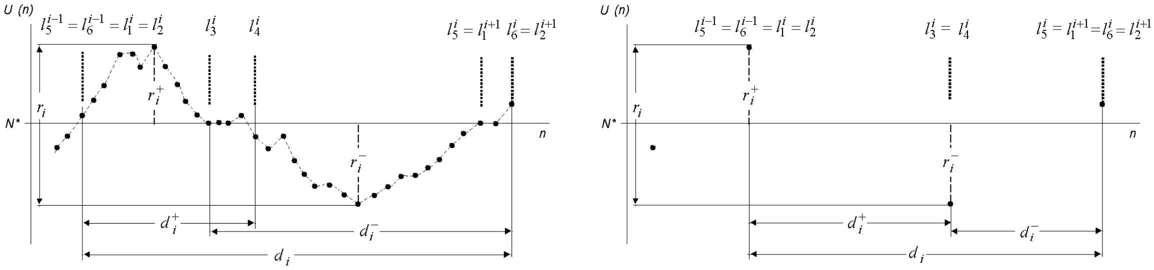
Исходные данные для решения поставленной задачи:

1. $U(n)$, $n=\overline{1,N}$ – образ звуковой волны (ОЗВ), или амплитудно-временное представление волнового процесса (АВП ВП).

2. Последовательность характеристик функции $U(n)$: границы квазипериодов (ПК) – $[l_1^i;l_6^i)$, где $i=\overline{1,M}$, M – количество ПК на АВП ВП; длины ПК – d_i; длины положительных и отрицательных полуволн ПК – d_i^+, d_i^-; амплитуды ПК, амплитуды положительных и отрицательных полуволн ПК $-r_i, r_i^+, r_i^-$; адреса локальных и глобальных экстремумов на ПК. ПК обозначим K_i, $i=\overline{1,M}$ (рис. 1). Следует отметить, что характеристики ПК получены с учетом двух уровней АЦП, которые мы отождествляем с уровнем нуля вольт. Рассуждения и метод определения границ ПК и его характеристик приведены в [8].

Необходимо:

– разработать метод автоматического определения границ текущих ОИ в ОЗВ, инвариантный к темпу, тембру и интенсивности принятого акустического сигнала. В качестве текущих объектов исследования должны выступать отрезки АВП ВП, соответствующие фонологическим единицам, в формировании которых участвует голосовой источник (для РС), и тональным сигналам (для специальных сигналов).



а) б)

Рисунок 1 – Характеристики K_i функции $U(n)$

Отметим, что представленная работа выполнена в рамках договора о творческом сотрудничестве между ИПММ НАН Украины (г. Донецк) и ИПМ им. М.В. Келдыша РАН (г. Москва, Россия).

1. Энергетически завершенные фрагменты в ОЗВ

Отказаться от проблемы использования «окон» при автоматическом выборе текущего ОИ возможно на основе следующих рассуждений.

Физический процесс в природе называют энергетически завершенным (ЭЗ), если его можно представить в виде следующих фаз его развития [11]: экскурсия (фаза зарождения – Э), выдержка (фаза существования – В) и рекурсия (фаза угасания – Р) (рис. 2).

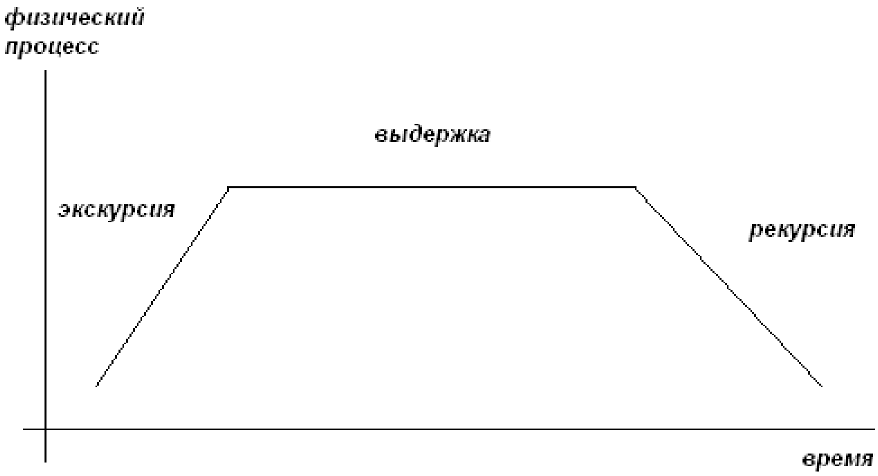


Рисунок 2 – Энергетически завершенный физический процесс

С физической точки зрения размах амплитуды ПК АВП ВП определяется энергией, которую несет в себе звуковая волна. Таким образом, задача состоит в локализации на АВП ВП объекта исследования, представляющего собой ЭЗ фрагмент (ЭЗФ) функции $U(n)$, изменений энергии на котором отображает указанные выше фазы.

В качестве примера можно привести изменение состояния органов голосового тракта, в результате которого происходит формирование элементарных звуковых форм, составляющих речевое сообщение. Очевидно, что для их реализации необходимо время и перестройка органов голосового тракта. Следовательно, структуру речевого сообщения можно рассматривать как последовательность элементарных звуковых форм, характеризующихся участками устойчивых состояний и переходных процессов (рис. 3 а, б).

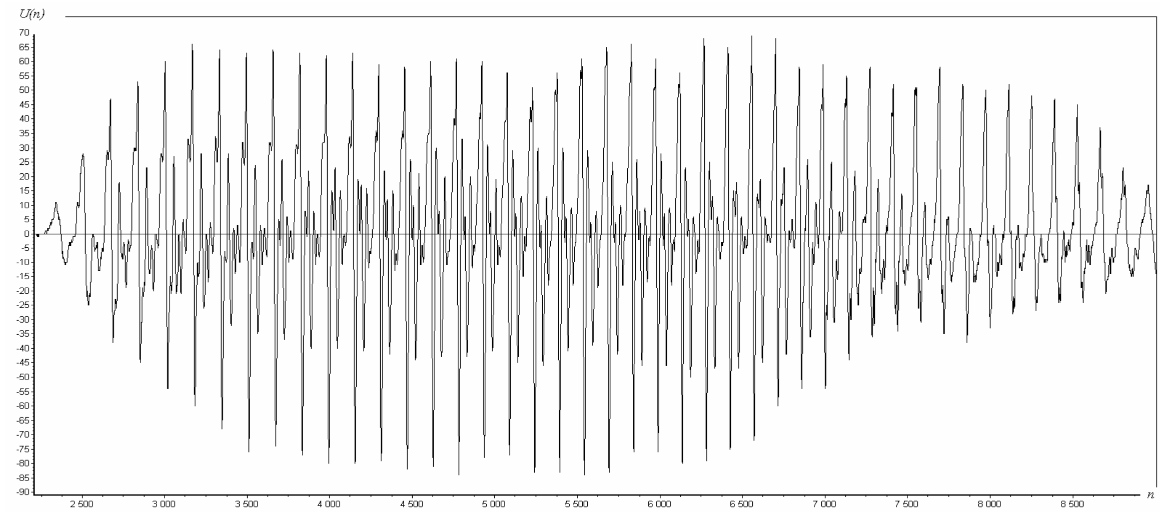
Акустический сигнал, полученный в результате функционирования дудочки (или свистка), также представляет собой последовательность элементарных звуковых форм, которые можно рассматривать в контексте завершенного физического процесса (рис. 3 в, г).

На основе анализа характера поведения последовательности характеристик $\{r_i\}_{i=1}^{M}$ на ПК $\{K_i\}_{i=1}^{M}$ определим ЭЗФ из условия:

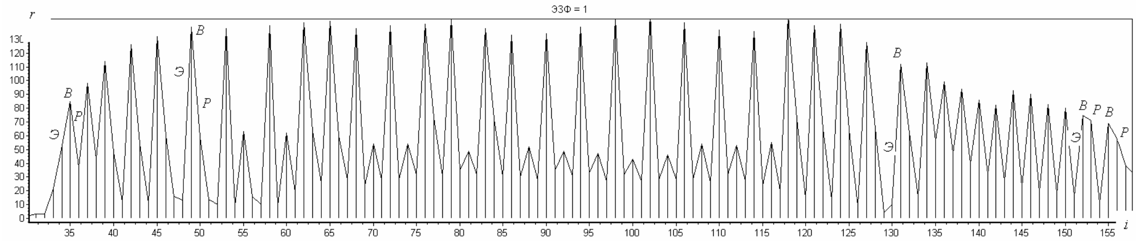
$$r_{i-1} > r_i = r_{i+p1},\ r_{i+p1} < r_{i+p1+1} < r_{i+p2-1} = r_{i+p2},\ r_{i+p2} > r_{i+p3-1} > r_{i+p3} = r_{i+p4} < r_{i+p4+1}, \quad (1)$$

где $s1 \in N : \forall p1 = \overline{0, s1}$, $r_i = r_{i+p1}$, $s2 \in N : \forall p2 = \overline{0, s2}$, $r_{i+p2-1} = r_{i+p2}$,

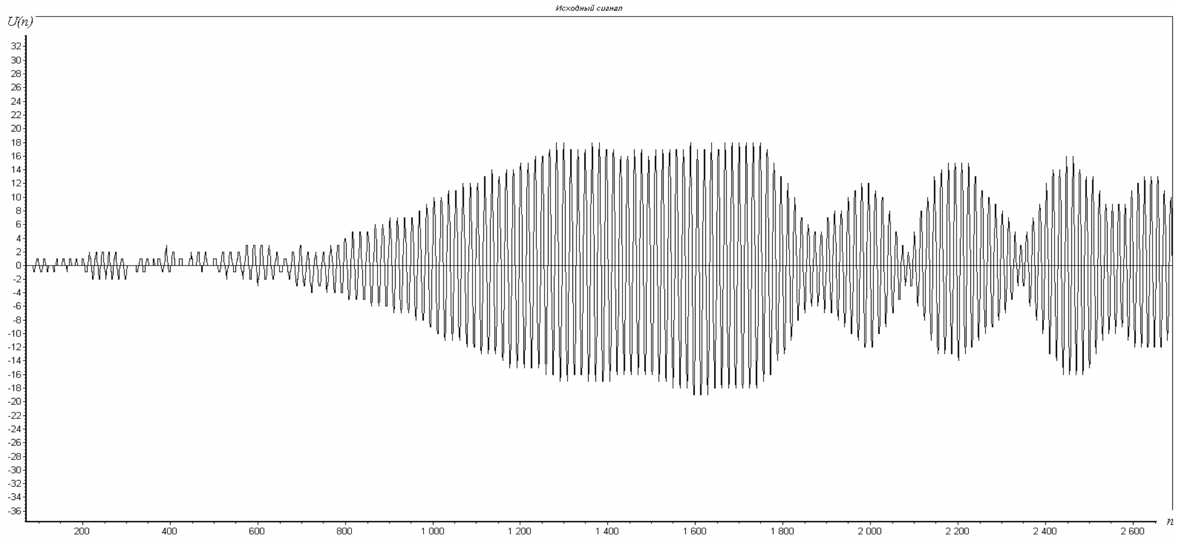
$s3 \in N : \forall p3 = \overline{0, s3}$, $r_{i+p3-1} > r_{i+p3}$, $s4 \in N : \forall p4 = \overline{0, s4}$, $r_{i+p3} = r_{i+p4}$.



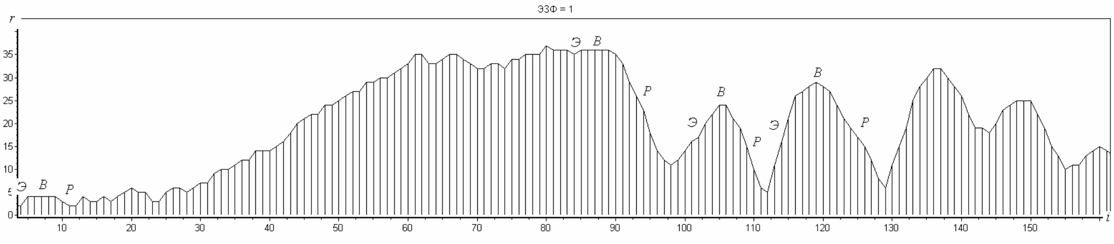
а)



б)



в)



г)

Рисунок 3 – а) АВП ВП фонемы «о», в) АВП ВП функционирование свистка, б) последовательность ЭЗФ для фонемы «о» и г) – для свистка

Границы ЭЗФ определяются аналогично границам ПК [9], [11].

$$lr_1^s : lr_1^s = l_1^i \quad для \quad r_{i-1} > r_i = r_{i+p1}, \tag{2}$$

$$lr_2^s : lr_2^s = l_6^{i+p1} \quad для \quad r_{i+p1} < r_{i+p1+1}, \tag{3}$$

$$lr_3^s : lr_2^s = l_6^{i+p2} \quad для \quad r_{i+p1+1} < r_{i+p2-1} = r_{i+p2} > r_{i+p3-1}, \tag{4}$$

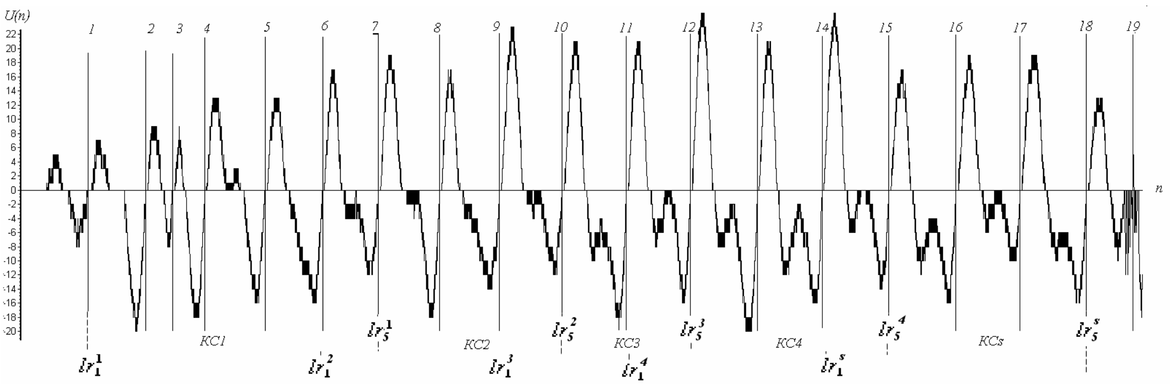
$$lr_4^s : lr_4^s = lr_1^{s+1} = l_1^{i+3} \quad для \quad r_{i+p3-1} > r_{i+p3} = r_{i+p4}, \tag{5}$$

$$lr_5^s : lr_5^s = lr_2^{s+1} = l_{61}^{i+p4} \quad для \quad r_{i+p3} = r_{i+p4} < r_{i+p4+1}. \tag{6}$$

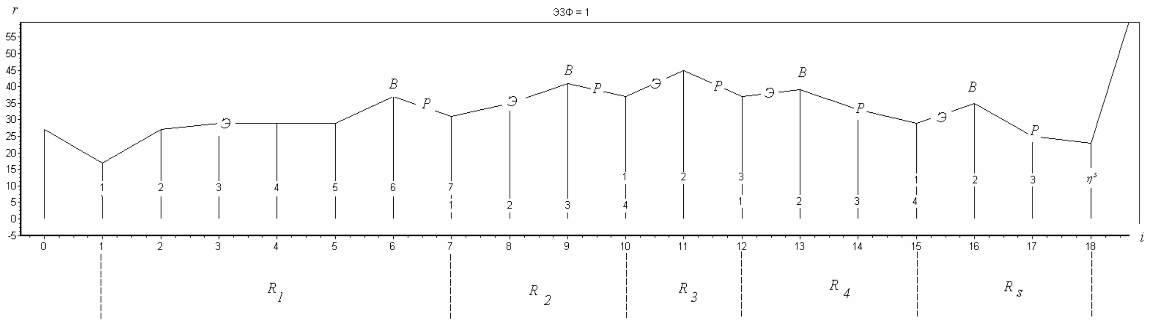
Отметим, что выдержка может быть представлена одной точкой.

На основании приведенных выше рассуждений дадим определение КС и ЭЗФ.

Определение 1. Полуинтервал $[lr_1^s; lr_5^s)$ функции $U(n)$, границы которого определены согласно соотношениям 1 – 6, будем называть квазипериодической структурой (КС) (рис. 4 а); последовательность характеристик ПК $\{r_\mu\}_{\mu=1}^{\eta^s}$ на этом полуинтервале, удовлетворяющим условию (1), будем называть энергетически завершенным фрагментом первого уровня; s-й ЭЗФ функции $U(n)$ обозначим R_s^1, η^s – количество ПК в R_s^1 (рис. 4 б).



а)



б)

Рисунок 4 – а) последовательность квазипериодических структур, согласованных с б) – последовательностью ЭЗ фрагментов

Введенное выше **определение 1** не позволяет решить задачу определения границ искомых ОИ, так как следует рассматривать последовательность ЭЗФ, а следовательно, последовательность КС, объединенных между собой на основе анализа «близости» характеристик АВП ВП на указанных фрагментах.

Анализ КС показал, что в рамках одного сигнала, излучаемых одним источником (диктором, свистком, дудочкой), их длины могут варьироваться, но должны быть близки. Характер изменения хотя бы одной характеристики на соответствующих КС должен быть одинаковым. Для решения задачи введем далее определения «близости» характеристик КС.

2. Однородные квазипериодические структуры

На основе представленного рассуждения следует, что границы текущих ОИ, содержащих КС, необходимо выбирать как последовательности *энергетически завершенных отделимых структур с близкими между собой характеристиками* без использования языковых критериев, безотносительно к фонологическому составу принятого речевого сообщения, а также информации о принятом АС.

Определение однородности характеристик ПК рассмотрено в [9], [13]. Дадим обобщенное определение «близости» двух фрагментов функции $U(n)$.

Определение 2. Сужение функции $U(n)$ на $[a;b)$ и сужение функции $U(n)$ на $[c;d)$ называют «близкими», если существует изоморфное отображение $F:\{(d_i^{'};r_i^{'})\}_{j=1}^{\xi} \xrightarrow{F} \{(d_i^{''};r_i^{''})\}_{j=1}^{\xi}$, где ξ – количество ПК на ЭЗФ с соответствующими характеристиками $\{(d_i^{'};r_i^{'})\}_{j=1}^{\xi}$ для $\underset{[a;b)}{U(n)}$ и $\{(d_i^{''};r_i^{''})\}_{j=1}^{\xi}$ для $\underset{[c;d)}{U(n)}$. «Близость» $\underset{[a;b)}{U(n)}$ и $\underset{[c;d)}{U(n)}$ будем обозначать $\underset{[a;b)}{U(n)} @ \underset{[c;d)}{U(n)}$.

Две КС, для которых выполнимо определение 2, будем называть однородными квазипериодическими структурами (ОКС) (рис. 5). ОКС и будут выступать в качестве текущих ОИ.

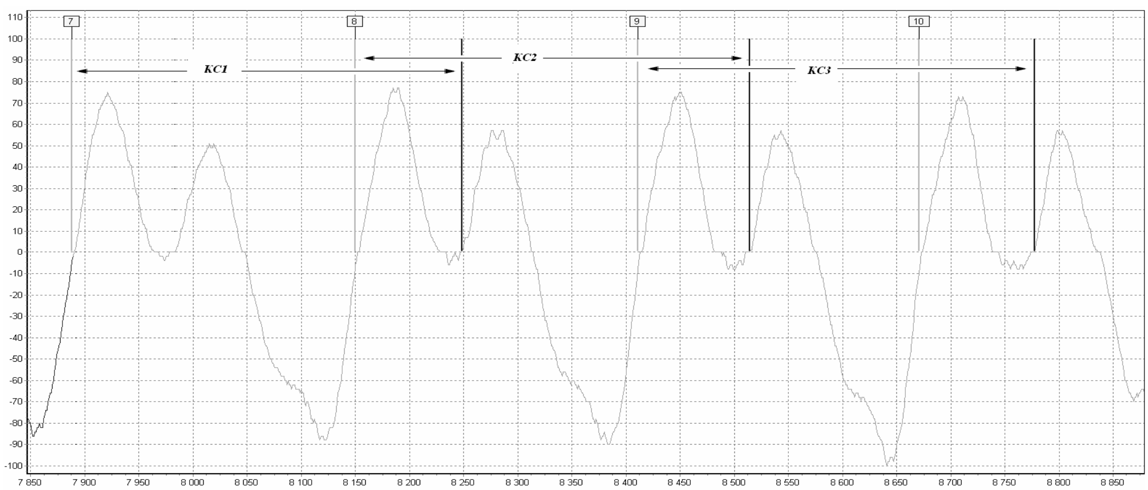


Рисунок 5 – Пример однородных квазипериодических структур

Изложенные выше рассуждения положены в основу алгоритма, реализующего автоматический выбор текущего ОИ, соответственно определение его границ в ОЗВ. Как было указано в постановке задачи, в качестве ОИ выступают ОКС, соответствующие голосовому источнику или ОТ.

3. Алгоритм автоматического определения границ текущих ОИ

На рис. 6 представлена структурная схема решения задачи автоматического определения границ текущих ОИ.

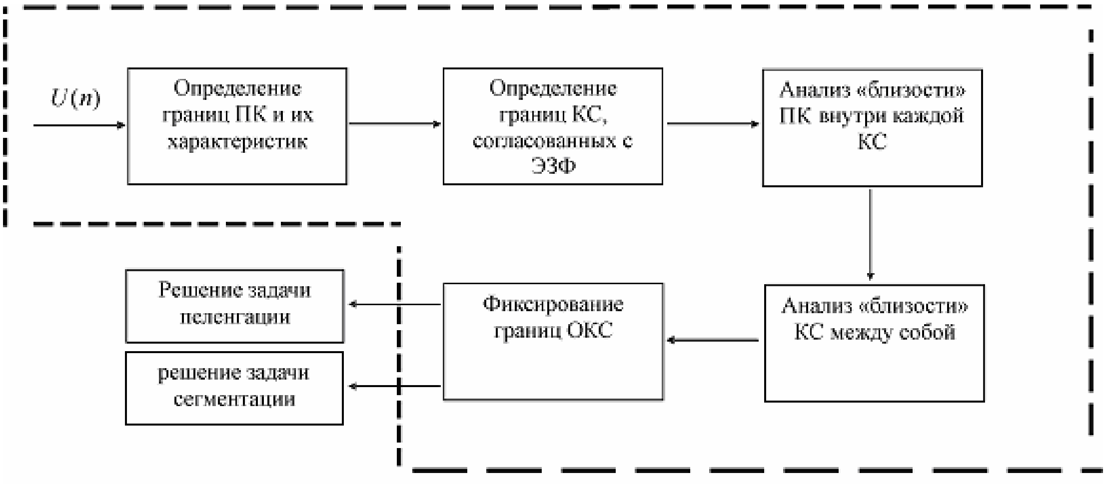


Рисунок 6 – Структурная схема алгоритма решения задачи автоматического определения границ текущих ОИ

Рассмотрим основные особенности его функционирования.

Как показано на рис. 6, алгоритм решения задачи автоматического определения границ текущих ОИ состоит из следующих блоков:

1) определение границ ПК и их характеристик – в данном блоке определяются границы ПК $[l_1^i;l_6^i)$ с учетом двух уровней АЦП, а также характеристики ПК – длины ПК $-d_i$; длины положительных и отрицательных полуволн ПК – d_i^+,d_i^-; амплитуды ПК, амплитуды положительных и отрицательных полуволн ПК – r_i,r_i^+,r_i^-; адреса локальных и глобальных экстремумов на ПК [9];

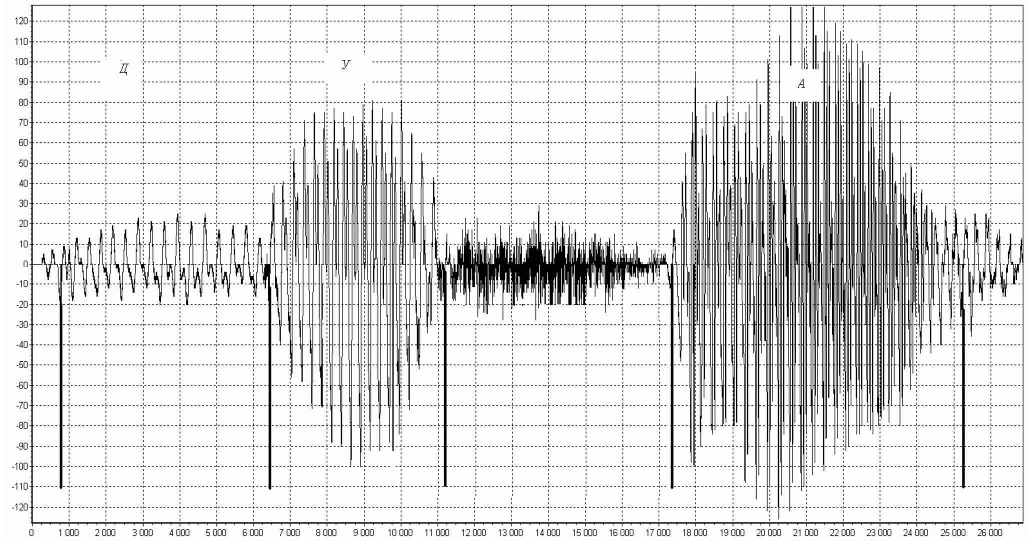
2) определение границ КС, согласованных с ЭЗФ – в этом блоке определяют ЭЗФ R_s^1 из условия 1 раздела 1, а также границы КС $[lr_1^s;lr_5^s)$, согласованные с ЭЗФ из условий 2 – 6 раздела 1;

3) анализ «близости» ПК внутри каждой КС – рассматриваются ПК в каждой КС, определяется количество похожих ПК. Правило определения близости в общем виде дано в разделе 2 (определение 2), подробное описание определения «близости» характеристик ПК рассмотрено в [9], [10];

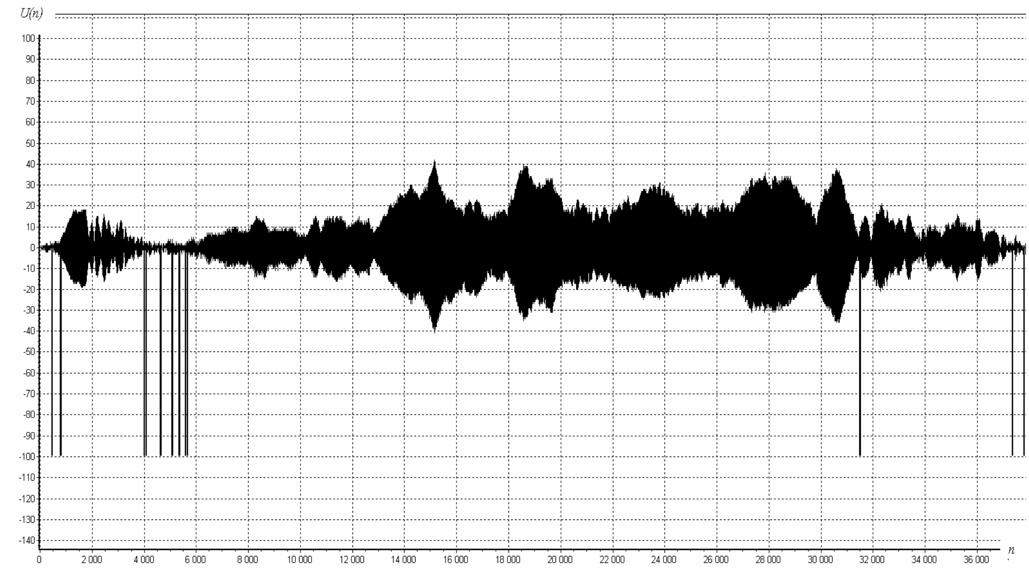
4) анализ близости КС между собой – после того, как рассмотрены все КС и определены близкие между собой ПК внутри каждой структуры, анализируются все найденные КС между собой на близость характеристик согласно определению 2. Отметим, что «похожие» КС между собой могут быть определены на некотором расстоянии друг от друга. Для такого случая КС должны иметь «пару». Если решение о «близости» КС не принято, то рассматриваются случаи «расщепления» сигнала – для этого последовательно объединяют слева (или справа) два ПК (для КС с большим количеством ПК η^s), далее определяют характеристики обобщенного ПК согласно блоку 1 и вновь производят анализ близости двух КС;

5) фиксирование границ ОКС – после полного анализа все КС принимается решение о фиксировании границ ОКС.

4. Экспериментальные исследования



а)



б)

Рисунок 7 – Пример работы алгоритма определения границ ОИ

Алгоритм определения ЭЗФ и ОКС был опробован на статистическом материале: для РС – на 13 дикторах (мужчин и женщин), каждый из которых произносил по 15 слов (список слов общий); для специальных сигналов – свисток и дудочка (длительность и амплитуда таких сигналов зависит от объема легких диктора и мощности воздуха, вдуваемого в отверстие свистка или дудочки).

На рис. 7 показан результат работы алгоритма определения границ ОИ: для РС – а), б) слова «душа», б) свисток.

Как показано на рис. 7 а), границы сегментов, соответствующие функционированию голосового источника, выделяются со 100 % точностью (выделены фонемы «д», «у», «а»), аналогичный результат получен для специальных сигналов (рис. 7 б).

Заключение

Основным результатом данной работы является разработка метода, на основании которого решены задачи выделения в образах звуковых волн однородных квазипериодических структур без константных и априорно-заданных величин. На основе предлагаемого метода разработан алгоритм, который успешно применен в задачах пеленгации на макетной установке СТС, а также при решении задач сегментации речевого сигнала – определение границ сегментов, соответствующих фонологическим единицам, в формировании которых участвует голосовой источник. Выводы относительно различных фрагментов АВП ВП производятся исключительно в результате анализа их характеристик.

Литература

1. Потапова Р.К. Речь: коммуникация, информация, кибернетика : учеб. пособие для вузов / Потапова Р.К. – М. : Радио и связь, 1997.
2. Pavlovsky V.E. Simulation and Experimental Elaboration of Acoustic Sensors for Mobile Robots / Vladimir E. Pavlovsky, Tatyana S. Khashan : paper presented at the RTO SET-092 Symposium on “Advanced Sensor Payloads for UAV”, (2-3 May, 2005).
3. Concept, Simulation and Elaboration of Audition Sensors for Robots / [Pavlovsky V.E., Myagkov A.S., Khashan T.S., Pavlovsky V.V.] : Proc. of The IARP Workshop “Adaptive and Intelligent Robots: Present and Future” (Moscow. – Nov. 24-26, 2005) / IPMech RAS, RAS. – Vol. 1. – P. 90-100.
4. Active audition for humanoid / [Nakadai K., Lourens T., Okuno H.G. and Kitano H.] : proc. of 17th National Conference on Artificial Intelligence (AAAI-2000).
5. Ермоленко Т.В. Использование непрерывного вейвлет-преобразования при распознавании вокализованных участков речевого сигнала / Т.В. Ермоленко // Искусственный интеллект. – 2004. – № 4. – С. 499-503.
6. Федоров Е.Е. Выделение длины периода основного тона речевого сигнала / Е.Е. Федоров // Искусственный интеллект. – 2004. – № 1. – С. 237-242.
7. Карпов О.Н. Технология построения устройств распознавания речи / Карпов О.Н. – Д. : Из-во Днепр. универс., 2001. – 184 с.
8. Шелепов В.Ю. К проблеме пофонемного распознавания / Шелепов В.Ю., Ниценко А.В. // Искусственный интеллект. – 2005. – № 4. – С. 662-668.
9. Хашан Т.С. Сегментация речевого сигнала / Т.С. Хашан // Искусственный интеллект. – 2007. – № 4. – С. 386-397.
10. Мышко С.В. Определение на амплитудно-временном представлении звуковых волн элементарных и однородных квазипериодических структур / С.В. Мышко, С.В. Блохин, А.В. Доценко // Биофизика сложных систем. – 1993. – С. 92-100.
11. Мышко Н.В. Определение энергетически завершенных фрагментов амплитудно-временного представления звуковых волн / Н.В. Мышко, А.В. Доценко, С.В. Блохин, С.В. Мышко // Биофизика сложных систем. – 1993. – С. 103-117.

Т.С. Хашан

Автоматичне визначення меж поточних об’єктів дослідження для рішення задач пеленгації, сегментації мовних і спеціальних сигналів

Розроблено метод автоматичного визначення меж поточних об’єктів дослідження в образах звукових хвиль, інваріантний до темпу, тембру й інтенсивності прийнятого акустичного сигналу. Поточними об’єктами дослідження виступають відрізки амиплітудно-часового подання хвильового процесу, що відповідають фонологічним одиницям, у формуванні яких бере участь голосове джерело (для мовних сигналів) і тональним сигналам (для спеціальних сигналів). На основі пропонованого методу розроблено алгоритм, що успішно застосовано у завданнях пеленгації на макетній установці системи технічного слуху.

T.S. Khashan

The Automated Determination of Boundaries of Current Object Investigations for Solving the Problems of Direction Finding, Segmentation of Speech and Special Signals

The method of extraction of quasi-periodic structures in images of sound waves is put in a basis of proposed algorithm of direction finding of an acoustic signal. Proposed method does not contain a priori known constants and is invariant to rate, timbre and intensity of the accepted acoustic signal. Presented results can be applied also for resolving the problem of segmentation of a speech signal.

Статья поступила в редакцию 03.02.2009.